

Finite-Length Multi-Input Multi-Output Channel Shortening Pre-filters

Related Application:

5 This application claims priority from Provisional application No. 60/158,713 filed on October 8, 1999. This application is also related to a Provisional application No. 60/158,714, also filed on October 8, 1999.

Background of the Invention:

10 The combination of maximum likelihood sequence estimation (MLSE) with receiver diversity is an effective technique for achieving high performance over noisy, frequency-selective, fading channels impaired by co-channel interference. With the addition of transmitter diversity, the resulting multi-input multi-output (MIMO) frequency-selective channel has a significantly higher capacity than its
15 single-input multi-output (SIMO) or single-input single-output (SISO) counterparts. The use of maximum likelihood multi-user detection techniques on these frequency-selective MIMO channels significantly outperforms single-user detection techniques that treat signals from other users as colored noise. However, MLSE complexity increases exponentially with the number of inputs (or transmit
20 antennas) and with the memory of the MIMO channel, making its implementation over severe inter-symbol interference (ISI) channels very costly.

 The MIMO channel can be modeled as a collection of FIR filters (i.e., an FIR filter between each input point (e.g., transmitting antenna) and each receiving point (e.g. receiving antenna), and the "memory of the channel" corresponds to the
25 number of taps in the FIR filters.

 The Discrete Matrix Multitone (DMMT) was shown to be a practical transceiver structure that asymptotically achieves the MIMO channel capacity when combined with powerful codes. It uses the Discrete Fourier Transform (DFT) to partition the frequency responses of the underlying frequency-selective
30 channels of the MIMO systems into a large number of parallel, independent, and (approximately) memoryless frequency subchannels. To eliminate inter-block and intra-block interference, a cyclic prefix whose length is equal to the MIMO channel

memory is inserted in every block. On severe-ISI MIMO channels, the cyclic prefix overhead reduces the achievable DMMT throughput significantly, unless a large FFT size is used which, in turn, increases the computational complexity, processing delay, and memory requirements in the receiver.

In short, the computation complexity increases exponentially with the number of taps in the FIR filters that may be used to model the channel.

N. Al-Dhahir and J. M. Cioffi, in "Efficiently-Computed Reduced-Parameter Input-Aided MMSE Equalizers for ML Detection: A Unified Approach," IEEE Trans. Information Theory, pp. 903-915, May 1996, disclose use of a time-domain pre-filter in the receiver to shorten the effective channel memory and hence reduce the cyclic prefix overhead and/or the number of MLSE states. The disclosed approach, however, is for SISO systems, and not for MIMO systems.

Summary

An advance in the art is realized with a MIMO pre-filter that is constructed from FIR filters with coefficients that are computed based on environment parameters that are designer-chosen. Given a transmission channel that is modeled as a set of FIR filters with memory ν , a matrix \mathbf{W} is computed for a pre-filter that results in an effective transmission channel \mathbf{B} with memory N_b , where $N_b < \nu$, where \mathbf{B} is optimized so that $\mathbf{B}_{opt} = \operatorname{argmin}_{\mathbf{B}} \operatorname{trace}(\mathbf{R}_{ee})$ subject to selected constraints; \mathbf{R}_{ee} being the error autocorrelation function. The coefficients of \mathbf{W} , which are sensitive to a variety of designer constraints, are computed by a processor within pre-filter at the front end of a receiver and loaded into an array of FIR filters that form the pre-filter.

Brief Description of the Drawing

FIG. 1 shows the major elements of a receiver in accord with the principles disclosed herein;

FIG. 2 presents the structure of pre-filter 30; and

FIG. 3 is a flowchart describing the method carried out by processor 220 within pre-filter 30.

Detailed Description

FIG. 1 depicts the general case of an arrangement with n_q transmitting antennas 11-1, 11-2, ... 11- n_q , that output signals (e.g., space-time encoded signals) to a transmission channel, and n_o receiving antennas 21-1, 21-2, ... 21- n_o . Each transmitting antenna p outputs a complex-valued signal x_p , the signals of the n_q antennas pass through a noisy transmission channel, and the n_o receiving antennas capture the signals that passed through the transmission channel. The received signals are oversampled by a factor of l in element 20 and applied to pre-filter 30. Thus, the sampling clock at the output of element 20 is of period $T_s = T/l$, where T is the inter-symbol period at the transmitting antennas. The transmission channel's characterization is also referenced to T_s . In the illustrative embodiment disclosed herein, therefore, pre-filter 20 develops n_i output signals that are applied to a conventional multi-input receiver 40, and the received signal can be expressed by

$$y_k^{(j)} = \sum_{i=1}^N \sum_{m=0}^{\nu^{(i,j)}} h_m^{(i,j)} x_{k-m}^{(i)} + n_k^{(j)}, \quad (1)$$

where $y_k^{(j)}$ is the signal at time k at the j^{th} receiving antenna, $h_m^{(i,j)}$ is the m^{th} coefficient (tap) in the channel impulse response between the i^{th} transmitting antenna and the j^{th} receiving antenna, and $n^{(j)}$ is the noise vector at the j^{th} receiving antenna. The memory of this path (i.e., the largest value of m for which $h_m^{(i,j)}$ is not zero) is denoted by $\nu^{(i,j)}$. It not unreasonable to assume, however, that the memory of the transmission channel is the same for all i,j links ($n_i \times n_o$ such links), in which case $\nu^{(i,j)} = \nu$. Alternatively, the $\nu^{(i,j)}$ limit in equation (1) can be set to that ν which corresponds to maximum length of all of the $n_i \times n_o$ channel input responses, i.e., $\nu = \max_{i,j} \nu^{(i,j)}$. All of these variables in equation (1) are actually $l \times 1$ column vectors, corresponding to the l time samples per symbol in the oversampled FIG. 1 arrangement. By grouping the received samples from all

n_o antennas at symbol time k into an $n_o \times 1$ column vector \mathbf{y}_k , one can relate \mathbf{y}_k to the corresponding $n_i \times 1$ (column) vector of input samples as follows

$$\mathbf{y}_k = \sum_{m=0}^v \mathbf{H}_m \mathbf{x}_{k-m} + \mathbf{n}_k, \quad (2)$$

where \mathbf{H}_m is the MIMO channel coefficients matrix of size $n_o \times n_i$, \mathbf{x}_{k-m} is a size $n_i \times 1$ input (column) vector, and \mathbf{n}_k is a size $n_o \times 1$ vector.

Over a block of N_f symbol periods, equation (2) can be expressed in matrix notation as follows:

$$\begin{bmatrix} \mathbf{y}_{k+N_f-1} \\ \mathbf{y}_{k+N_f-2} \\ \vdots \\ \mathbf{y}_k \end{bmatrix} = \begin{bmatrix} \mathbf{H}_0 & \mathbf{H}_1 & \cdots & \mathbf{H}_v & 0 & \cdots & 0 \\ 0 & \mathbf{H}_0 & \mathbf{H}_1 & \cdots & \mathbf{H}_v & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \mathbf{H}_0 & \mathbf{H}_1 & \cdots & \mathbf{H}_v \end{bmatrix} \begin{bmatrix} \mathbf{x}_{k+N_f-1} \\ \mathbf{x}_{k+N_f-2} \\ \vdots \\ \mathbf{x}_{k-v} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_{k+N_f-1} \\ \mathbf{n}_{k+N_f-2} \\ \vdots \\ \mathbf{n}_k \end{bmatrix} \quad (3)$$

or, more compactly,

$$\mathbf{y}_{k+N_f-1:k} = \mathbf{H} \mathbf{x}_{k+N_f-1:k-v} + \mathbf{n}_{k+N_f-1:k}. \quad (4)$$

The subscripts in equation (4) indicate a range. For example $k + N_f - 1 : k$ indicates the range $k + N_f - 1$ and k , inclusive.

It is useful to define the following correlation matrices:

$$\mathbf{R}_{xy} \equiv E[\mathbf{x}_{k+N_f-1:k-v} \mathbf{y}_{k+N_f-1:k}^*] = \mathbf{R}_{xx} \mathbf{H}^* \quad (5)$$

$$\mathbf{R}_{yy} \equiv E[\mathbf{y}_{k+N_f-1:k} \mathbf{y}_{k+N_f-1:k}^*] = \mathbf{H} \mathbf{R}_{xx} \mathbf{H}^* + \mathbf{R}_{nn}, \quad (6)$$

$$\mathbf{R}_{xx} \equiv E[\mathbf{x}_{k+N_f-1:k-v} \mathbf{x}_{k+N_f-1:k-v}^*] \text{ and} \quad (7)$$

$$\mathbf{R}_{nn} \equiv E[\mathbf{n}_{k+N_f-1:k} \mathbf{n}_{k+N_f-1:k}^*]. \quad (8)$$

It is assumed that these correlation matrices do not change significantly in time or, at least, do not change significantly over a time interval that corresponds to a TDMA burst (assumed to be much shorter than the channel coherence time), which is much longer than the length of the pre-filter, in symbol periods denoted by N_f . Accordingly, a re-computation of the above matrices, and the other parameters disclosed herein, leading to the computation of pre-filter coefficients, need not take place more often than once every TDMA burst.

Once \mathbf{H} , \mathbf{R}_{xx} and \mathbf{R}_{nn} and known, \mathbf{R}_{xy} and \mathbf{R}_{yy} are computed by $\mathbf{R}_{xx}\mathbf{H}^*$ and $\mathbf{H}\mathbf{R}_{xx}\mathbf{H}^* + \mathbf{R}_{nn}$, respectively.

Given the MIMO channel matrix \mathbf{H} with $\nu + 1$ members ($\mathbf{H}_0, \mathbf{H}_1, \dots, \mathbf{H}_\nu$), the objective is to create a MIMO pre-filter \mathbf{W} (element 30 in FIG. 1) with N_f matrix taps, i.e., matrix $\mathbf{W} \equiv [\mathbf{W}_0 \ \mathbf{W}_1 \ \dots \ \mathbf{W}_{N_f-1}]^T$, that equalizes \mathbf{H} so as to create an overall transmission channel for receiver 40 that corresponds to a matrix \mathbf{B} with memory N_b , where $N_b \ll \nu$.

The matrix \mathbf{B} can be expressed as $\mathbf{B} \equiv [\mathbf{B}_0 \ \mathbf{B}_1 \ \dots \ \mathbf{B}_{N_b}]^T$ where each \mathbf{B}_i is of size $n_i \times n_i$.

The MIMO channel-shortening pre-filter \mathbf{W} (element 30) is conditioned, or adjusted, to minimize the equalization Mean Squared Error (MSE), defined by $MSE \equiv \text{trace}(\mathbf{R}_{ee})$, where \mathbf{R}_{ee} is the autocorrelation matrix of the error vector \mathbf{E}_k that is given by

$$\mathbf{E}_k = \tilde{\mathbf{B}}^* \mathbf{x}_{k+N_f-1:k-\nu} - \mathbf{W}^* \mathbf{y}_{k+N_f-1:k}, \quad (9)$$

where the augmented MIMO matrix, $\tilde{\mathbf{B}}^*$, is

$$\tilde{\mathbf{B}}^* \equiv \begin{bmatrix} 0_{n_i \times n_i \Delta} & \mathbf{B}_0^* & \mathbf{B}_1^* & \dots & \mathbf{B}_{N_b}^* & 0_{n_i \times n_i s} \end{bmatrix} \equiv \begin{bmatrix} 0_{n_i \times n_i \Delta} & \mathbf{B}^* & 0_{n_i \times n_i s} \end{bmatrix}, \quad (10)$$

Δ is the decision delay that lies in the range $0 \leq (N_f + \nu - N_b - 1)$, and

$s \equiv N_f + \nu - N_b - \Delta - 1$. The $n_i \times n_i$ error autocorrelation function \mathbf{R}_{ee} can be expressed by the following:

$$\begin{aligned} \mathbf{R}_{ee} &\equiv E[\mathbf{E}_k \mathbf{E}_k^*] \\ &= \tilde{\mathbf{B}}^* (\mathbf{R}_{xx} - \mathbf{R}_{xy} \mathbf{R}_{yy}^{-1} \mathbf{R}_{yx}) \tilde{\mathbf{B}} \\ &= \tilde{\mathbf{B}}^* \mathbf{R}^\perp \tilde{\mathbf{B}} \\ &= \tilde{\mathbf{B}}^* \bar{\mathbf{R}} \mathbf{B}, \end{aligned} \quad (11)$$

where $\bar{\mathbf{R}}$ is a sub-matrix of \mathbf{R}^\perp determined by Δ .

Using the orthogonality principle, which states that $E[\mathbf{E}_k \mathbf{y}_{k+N_f-1:k}^*] = 0$ it can be shown that the optimum channel-shortening pre-filter and target impulse response filters (\mathbf{W} and \mathbf{B} , respectively) are related by

$$\begin{aligned}
\mathbf{W}_{opt}^* &= \tilde{\mathbf{B}}_{opt}^* \mathbf{R}_{xy} \mathbf{R}_{yy}^{-1} \\
&= \tilde{\mathbf{B}}_{opt}^* \mathbf{R}_{xx} \mathbf{H}^* (\mathbf{H} \mathbf{R}_{xx} \mathbf{H}^* + \mathbf{R}_{nn})^{-1} \\
&= \tilde{\mathbf{B}}_{opt}^* (\mathbf{R}_{xx}^{-1} + \mathbf{H}^* \mathbf{R}_{nn}^{-1} \mathbf{H})^{-1} \mathbf{H}^* \mathbf{R}_{nn}^{-1}.
\end{aligned} \tag{12}$$

The last line shows explicitly that the MIMO channel-shortening pre-filter consists of a noise whitener \mathbf{R}_{nn}^{-1} , a MIMO matched filter \mathbf{H}^* , and a bank of FIR channel-shortening pre-filter elements.

5 It remains to optimize $\tilde{\mathbf{B}}$ such that the MSE is minimized, which may be obtained by computing the parameters of \mathbf{B} that, responsive to specified conditions, minimizes the trace (or determinant) of \mathbf{R}_{ee} . The following discloses two approaches to such optimization.

Under one optimization approach the coefficients of \mathbf{B} are constrained so that some coefficient of \mathbf{B} is equal to the identity matrix, \mathbf{I} . A solution subject to this Identity Tap Constraint (ITC) can be expressed by

$$\mathbf{B}_{opt}^{ITC} = \operatorname{argmin}_B \operatorname{trace}(\mathbf{R}_{ee}) \text{ subject to } \mathbf{B}^* \Phi = \mathbf{I}_{n_i}, \tag{13}$$

where $\Phi^* \equiv \begin{bmatrix} \mathbf{0}_{n_i \times n_i m} & \mathbf{I}_{n_i} & \mathbf{0}_{n_i \times n_i (N_b - m)} \end{bmatrix}$ and $0 \leq m \leq N_b$. It can be shown that the optimum MIMO target impulse response and the corresponding error autocorrelation matrix are given by

$$\mathbf{B}_{opt}^{ITC} = \bar{\mathbf{R}}^{-1} \Phi (\Phi^* \bar{\mathbf{R}}^{-1} \Phi)^{-1} \text{ and} \tag{14}$$

$$\mathbf{R}_{ee, \min}^{ITC} = (\Phi^* \bar{\mathbf{R}}^{-1} \Phi)^{-1}. \tag{15}$$

As indicated above, $\bar{\mathbf{R}}$ is affected by the delay parameter Δ . Unless dictated by the designer, the delay parameter Δ , which can range between 0 and

20 $(N_f + \nu - N_b - 1)$, is chosen to minimize the trace of $\mathbf{R}_{ee, \min}^{ITC}$. Similarly, the index parameter m , which ranges between 0 and N_b , and which that affects Φ , is chosen to minimize the trace of $\mathbf{R}_{ee, \min}^{ITC}$.

Under a second optimization approach the imposed constraint is $\mathbf{B}^* \mathbf{B} = \mathbf{I}_{n_i}$. A solution subject to this Ortho-Normality Constraint (ONC) can be expressed by

$$25 \quad \mathbf{B}_{opt}^{ONC} = \operatorname{argmin}_B \operatorname{trace}(\mathbf{R}_{ee}) \text{ subject to } \mathbf{B}^* \mathbf{B} = \mathbf{I}_{n_i}, \tag{16}$$

Defining the eigen-decomposition

$$\bar{\mathbf{R}} \equiv \mathbf{U}\mathbf{\Sigma}\mathbf{U}^* = \mathbf{U} \text{diag}(\sigma_0, \sigma_1 \cdots \sigma_{n_i(N_b+1)-1}) \mathbf{U}^*, \quad (17)$$

where $\sigma_0 \geq \sigma_1 \cdots \geq \sigma_{n_i(N_b+1)-1}$, then the optimum MIMO target response and the resulting error autocorrelation matrix are given by

$$\mathbf{B}_{opt}^{ONC} = \mathbf{U} \begin{bmatrix} \mathbf{e}_{n_i N_b} & \cdots & \mathbf{e}_{n_i(N_b+1)-1} \end{bmatrix}, \quad (18)$$

5 where \mathbf{e}_i is unit vector with a 1 at position i , and 0's elsewhere, and

$$\mathbf{R}_{ee,min}^{ONC} = \text{diag}(\sigma_{n_i N_b}, \cdots, \sigma_{n_i(N_b+1)-1}). \quad (19)$$

Illustratively, if $n_i=3$ and $N_b=3$, $\mathbf{B}_{opt}^{ONC} = \mathbf{U}[\mathbf{e}_9, \mathbf{e}_{10}, \mathbf{e}_{11}]$, meaning that \mathbf{B}_{opt}^{ONC} is a three column matrix comprising the 9th through the 11th columns of matrix \mathbf{U} . Stated in words, the optimum MIMO target impulse response matrix is given by the n_i eigenvectors of $\bar{\mathbf{R}}$ that correspond to its n_i smallest eigenvalues. The delay parameter Δ ($0 \leq \Delta \leq N_f + \nu - N_b - 1$) that affects $\bar{\mathbf{R}}$ is optimized to minimize the trace (or determinant) of $\mathbf{R}_{ee,min}^{ONC}$.

With the above analysis in mind, a design of a prefilter 30 can proceed for any given set of system parameters, which includes:

- MIMO channel memory between the input points and the output point of the actual transmission channel, ν ,
- The number of pre-filter taps chosen, N_f ,
- The shortened MIMO memory, N_b ,
- The number of inputs to the transmission channel, n_i ,
- The number of output derived from the transmission channel, n_o ,
- The autocorrelation matrix of the inputs, \mathbf{R}_{xx} ,
- The autocorrelation matrix of the noise, \mathbf{R}_{nn} ,
- The oversampling used, l , and
- The decision delay, Δ .

25 The structure of pre-filter 30 is shown in FIG. 2. In the illustrated embodiment, it comprises two main components: processor 220 and filter section 210.

Filter section 210 in the FIG. 2 illustrative example comprises a collection of FIR filters that connect the n_o input array of signals from sampling circuit 20 to an

n_i output array of points. That is, there are $j \times i$ FIR filters $P_{j,i}$, that couple input point j to output point i .

Processor 220 is responsive to the n_o signals received by antennas 21 and sampled by circuit 20, and it computes the coefficients of W , as disclosed above.

- 5 W_0 is a matrix that defines the coefficients in the 0th tap of the $j \times i$ FIR filters, W_1 is a matrix that defines the coefficients in the 1st tap of the $j \times i$ FIR filters, etc.

- The method of developing the parameters of pre-filter 30, carried out in processor 220, is shown in FIG. 3. Block 100 develops an estimate of the MIMO channel between the input points and the output point of the actual transmission channel. This is accomplished in a conventional manner through the use of training sequences. The estimate of the MIMO channel can be chosen to be limited to a given memory length, ν , or can be allowed to include as much memory as necessary to reach a selected estimate error level. That, in turn, depends on the environment and is basically equal to the delay spread divided by T_s .

Following step 100, step 110 determines the matrices, \mathbf{R}_{nn} , \mathbf{R}_{xx} , \mathbf{R}_{xy} , and \mathbf{R}_{yy} . The matrix \mathbf{R}_{nn} is computed by first computing $\mathbf{n} = \mathbf{y} - \mathbf{H}\mathbf{x}$ and then computing the expected value $E[\mathbf{n}^* \mathbf{n}]$ -- see equation (8) above. The matrix \mathbf{R}_{xx} is computed from the known training sequences -- see equation (7) above -- (or is pre-computed and installed in processor 220). It may be noted that for uncorrelated inputs, $\mathbf{R}_{xx} = \mathbf{I}$. The matrices \mathbf{R}_{xy} and \mathbf{R}_{yy} are computed from the known training sequences and the received signal or directly from \mathbf{H} and \mathbf{R}_{nn} -- see equations (5) and (6) above.

- Following step 110, step 120 computes $\mathbf{R}^\perp = \mathbf{R}_{xx} - \mathbf{R}_{xy} \mathbf{R}_{yy}^{-1} \mathbf{R}_{yx}$, and the sub-matrix $\bar{\mathbf{R}}$. From equation (10) it can be seen that $\bar{\mathbf{R}}$ is obtained by dropping the first $n_i \Delta$ rows and the last $n_i s$ rows of \mathbf{R}^\perp .

In accordance with the ITC approach, selecting some value of $0 \leq m \leq N_b$ allows completion of the design process. Accordingly, following step 120, step 130 chooses a value for m , develops $\Phi^* \equiv \begin{bmatrix} \mathbf{0}_{n_i \times n_i m} & \mathbf{I}_{n_i} & \mathbf{0}_{n_i \times n_i (N_b - m)} \end{bmatrix}$ and carries out

the computation of equation (13). Step 140 finally develops the coefficients of matrix W in accordance with equation (12), and installs the developed coefficients within filter 210.

In accordance with the ONC approach, step 130 computes the matrix U in a conventional manner, identifies the unit vectors e_i , and thus obtains the matrix B . Step 140 develops the coefficients of matrix W in accordance with equation (12), and installs the developed coefficients within filter 210.

It should be understood that a number of aspects of the above disclosure, for example, those related to the ITC constraint and to the ONC constraint, are merely illustrative, and that persons skilled in the art may make various modifications that, nevertheless, are within the spirit and scope of this invention. For example, the pre-filter described above generates a multi-output signal, with the number of outputs being n_i , that being also the number of transmitting antennas 11. This, however, is not a limitation of the principles disclosed herein. The number of pre-filter outputs can, for example, be larger than n_i , for example as high as $n_i(N_b + 1)$. The performance of the receiver will be better with more filter outputs, but more outputs require more FIR filters, more FIR filter coefficients, and correspondingly, a greater processing power requirement placed on processor 220.